Using classification trees to identify bumble bees

Rebekah Brassfield
rb164655@umconnect.umt.edu

Janene Lichtenberg Mrs
Salish Kootenai College, janene_lichtenberg@skc.edu

Using Classification Trees to Identify Bumble Bees

Rebekah Brassfield

Identification of Bumble Bees

Identification of bumble bees is difficult without a microscope.

This requires lethally collecting specimens for ID.

Identification of Bumble Bees

Recently, high resolution cameras have been used to photograph key features, so we do not have to collect individuals

Face Length

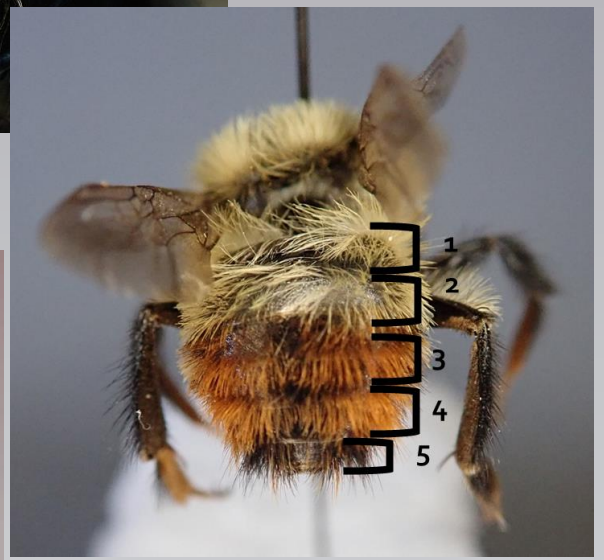
Short, medium, long, cuckoo

Face Color

Yellow, black, white, mixed

Abdominal Pattern

5 segment colors



Thorax Pattern

Stripe, spot, notch

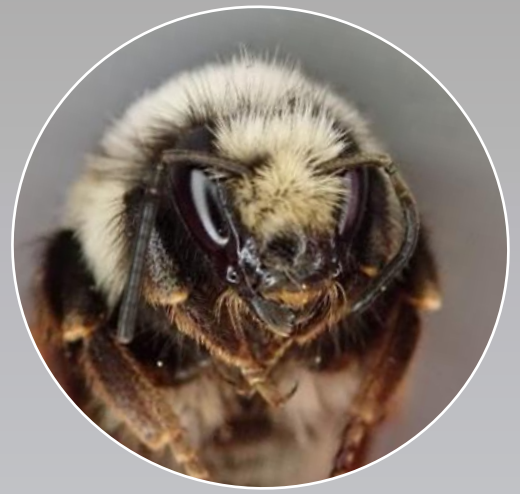
Let’s say we want to identify this bee:

Side view

Top view

Face View

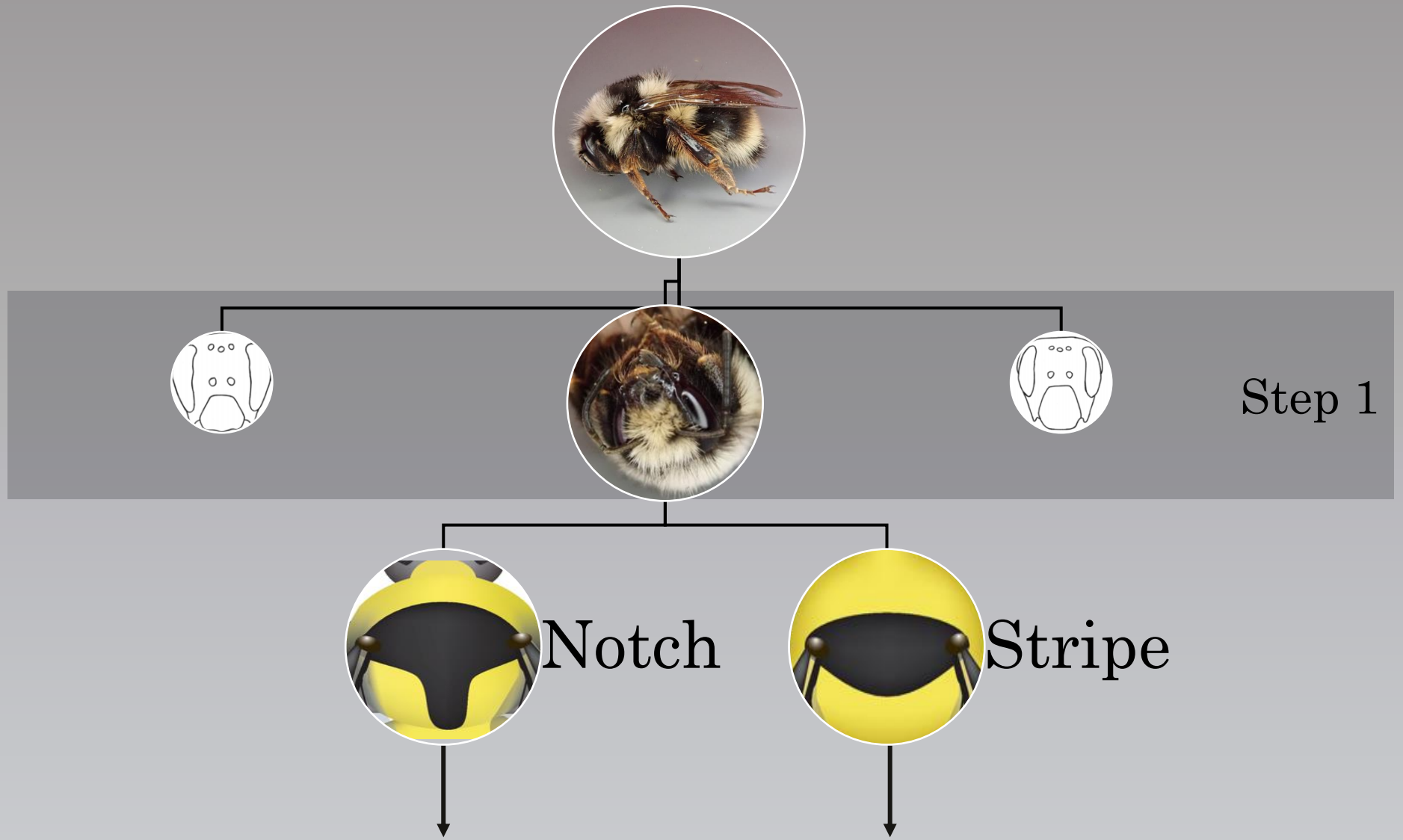
Does it have a short, medium or long face?



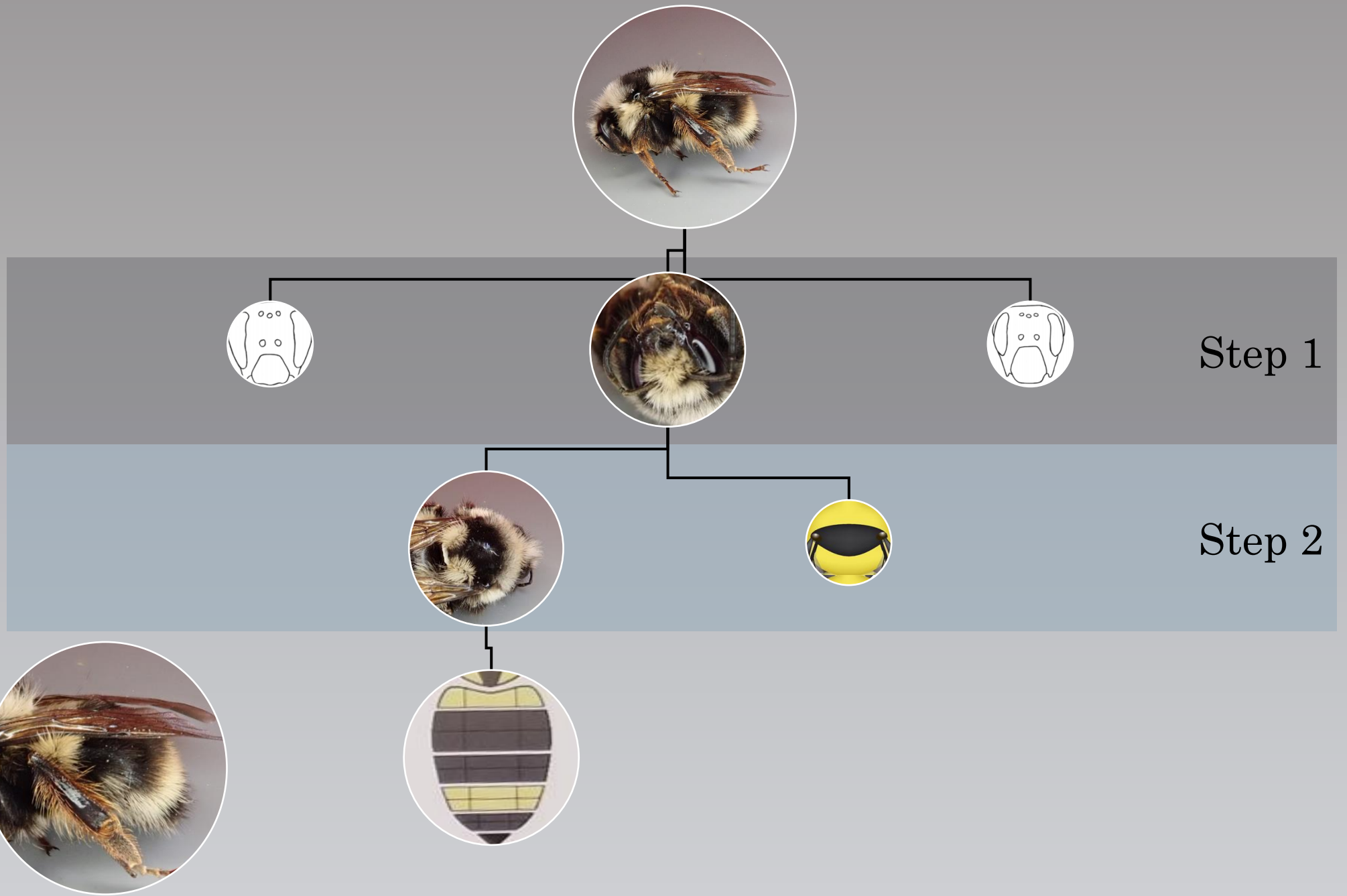
Short

Medium

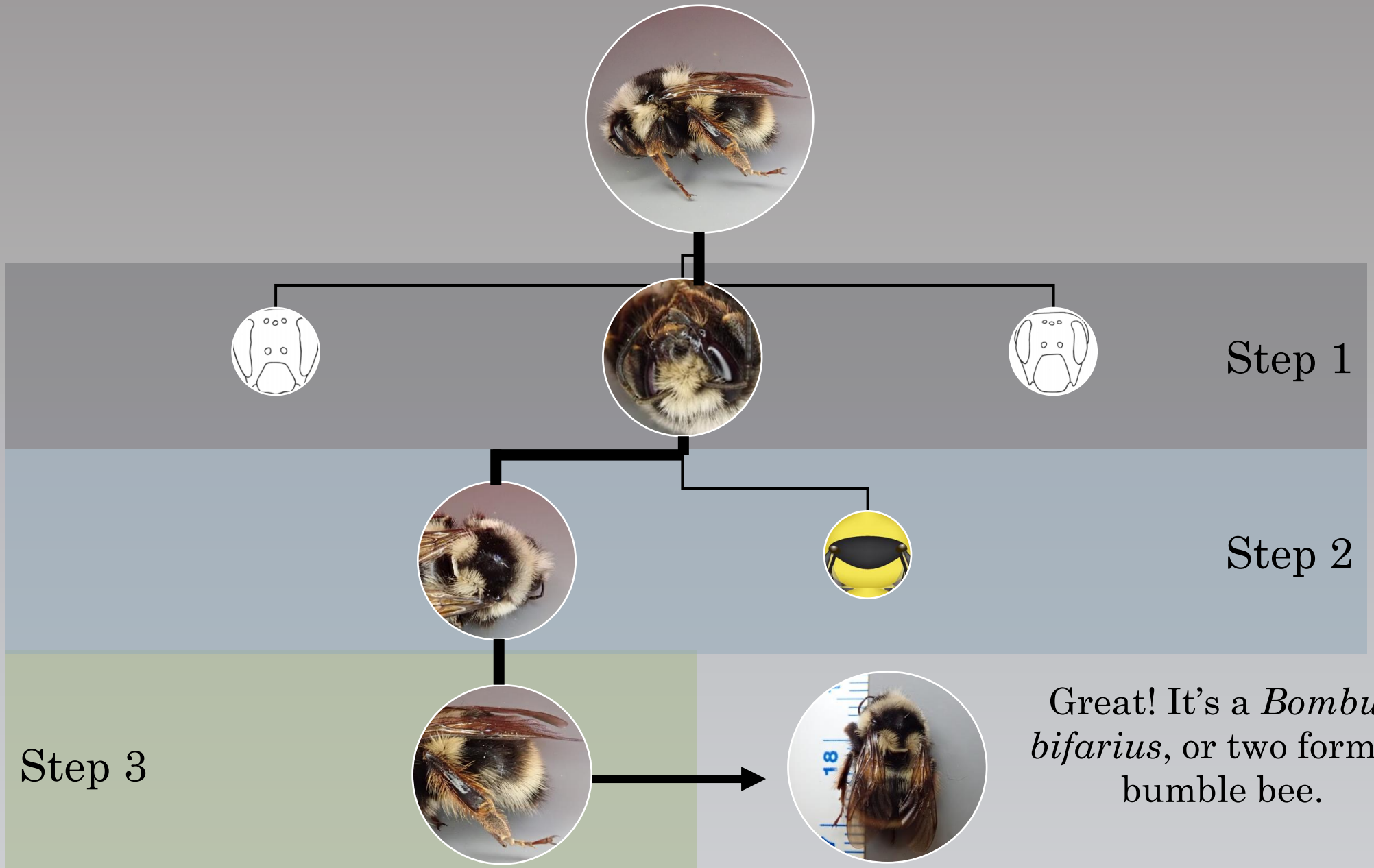
Long



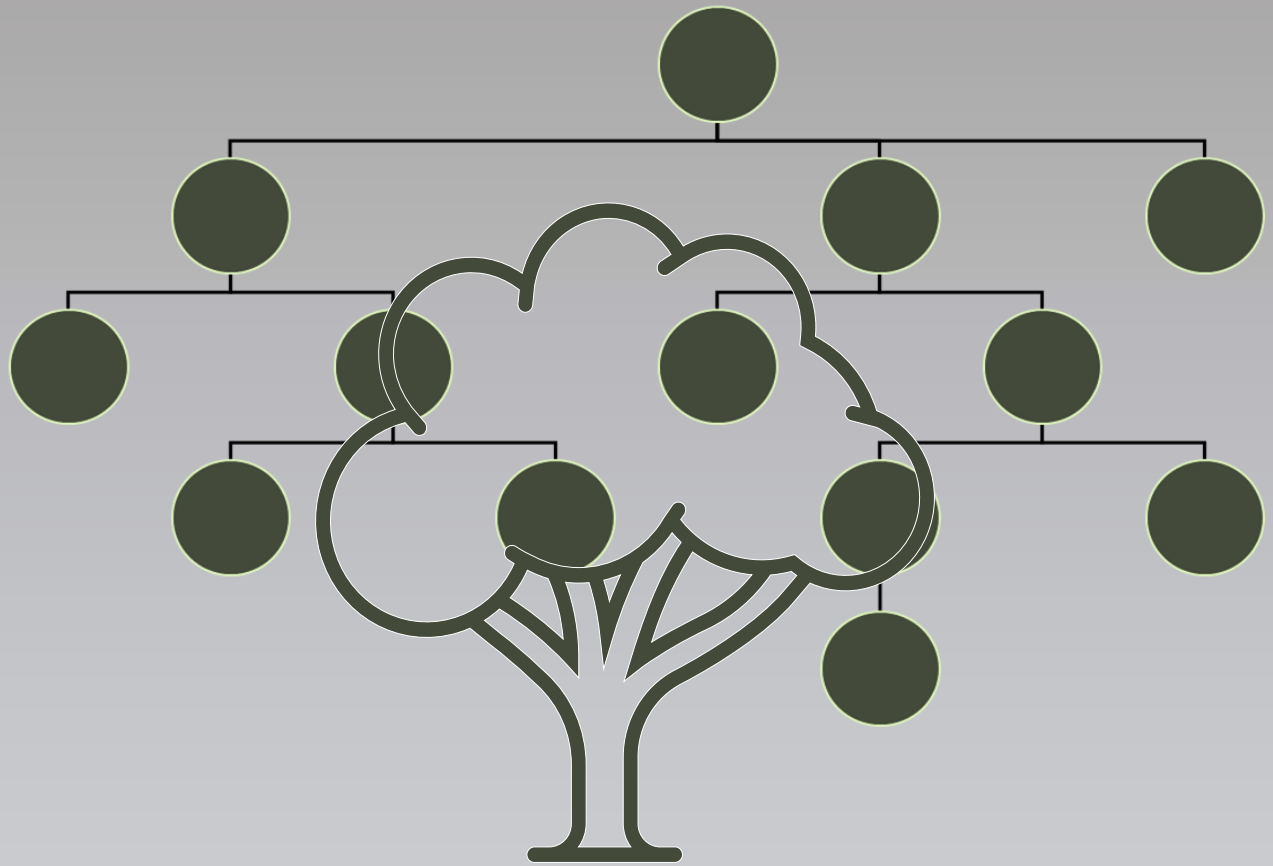
Does it have a notch or a stripe?



Is the abdominal pattern YBBYB?

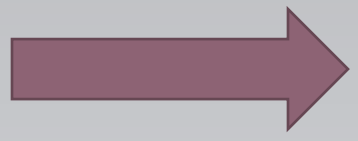

Step 1
Step 2
Step 3
Great! It's a *Bombus bifarius*, or two formed bumble bee.

This is a classification tree…



A machine learning technique in which a classification tree goes from observations about an item, to a conclusion about the item's target value.

Observations	
Face Length:	Short, Medium, Long, Cuckoo
Face Color:	Yellow, Black, White, Mixed
Thorax Pattern:	Stripe, Spot, Notch
Abdomen Pattern:	5 segment colors: 1 2 3 4 5



Target Value

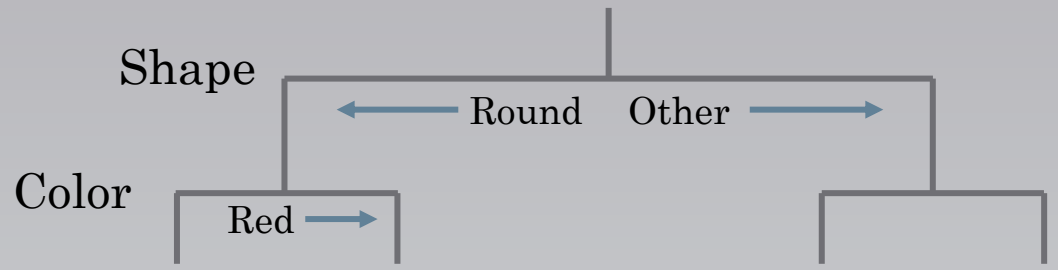
Species:

appositus	insularius
bifarius	melanopygus
centralis	mixtus
fervidus	vagans
flavifrons	

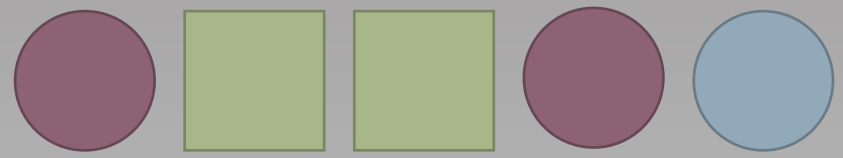
Two classification tree models

1. CART

Grows a single tree that branches based on the *majority* of the data

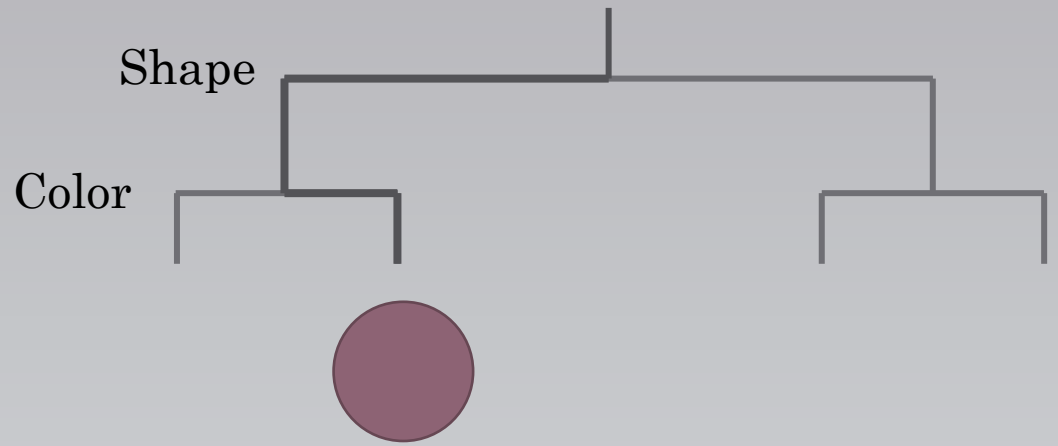


1. CART



And stops branching when the values are matching

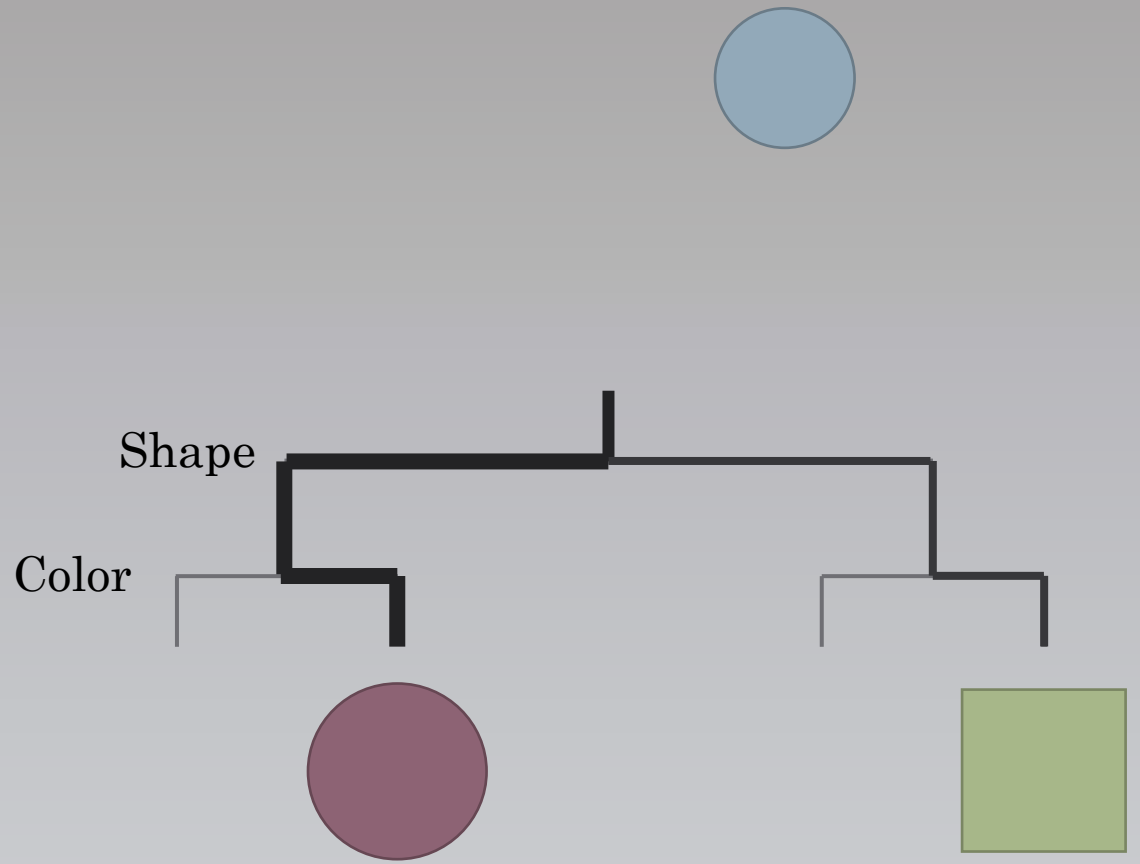
The data dictates the shape of the tree



1. CART

So over time it becomes more accurate with values it encounters more often

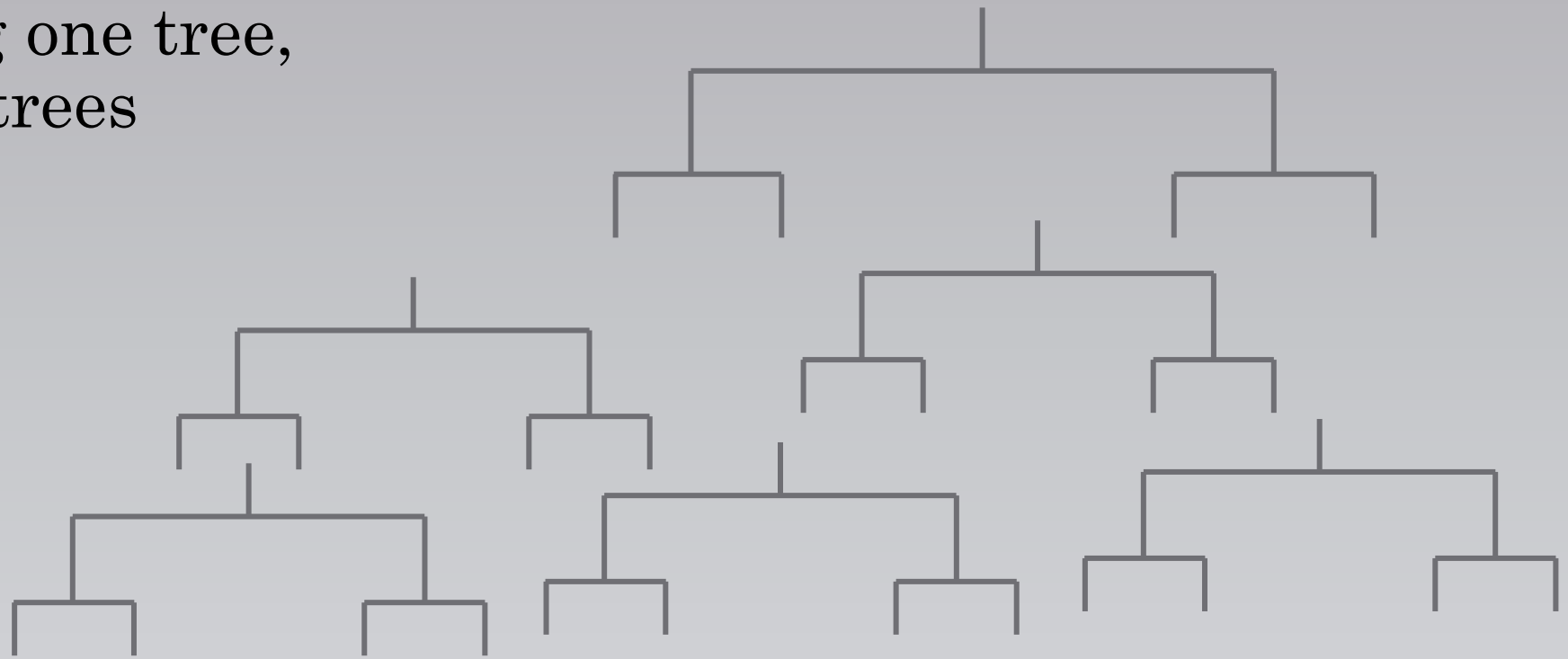
…which can lead to errors



Two classification tree models

2. Random Forest (RF)

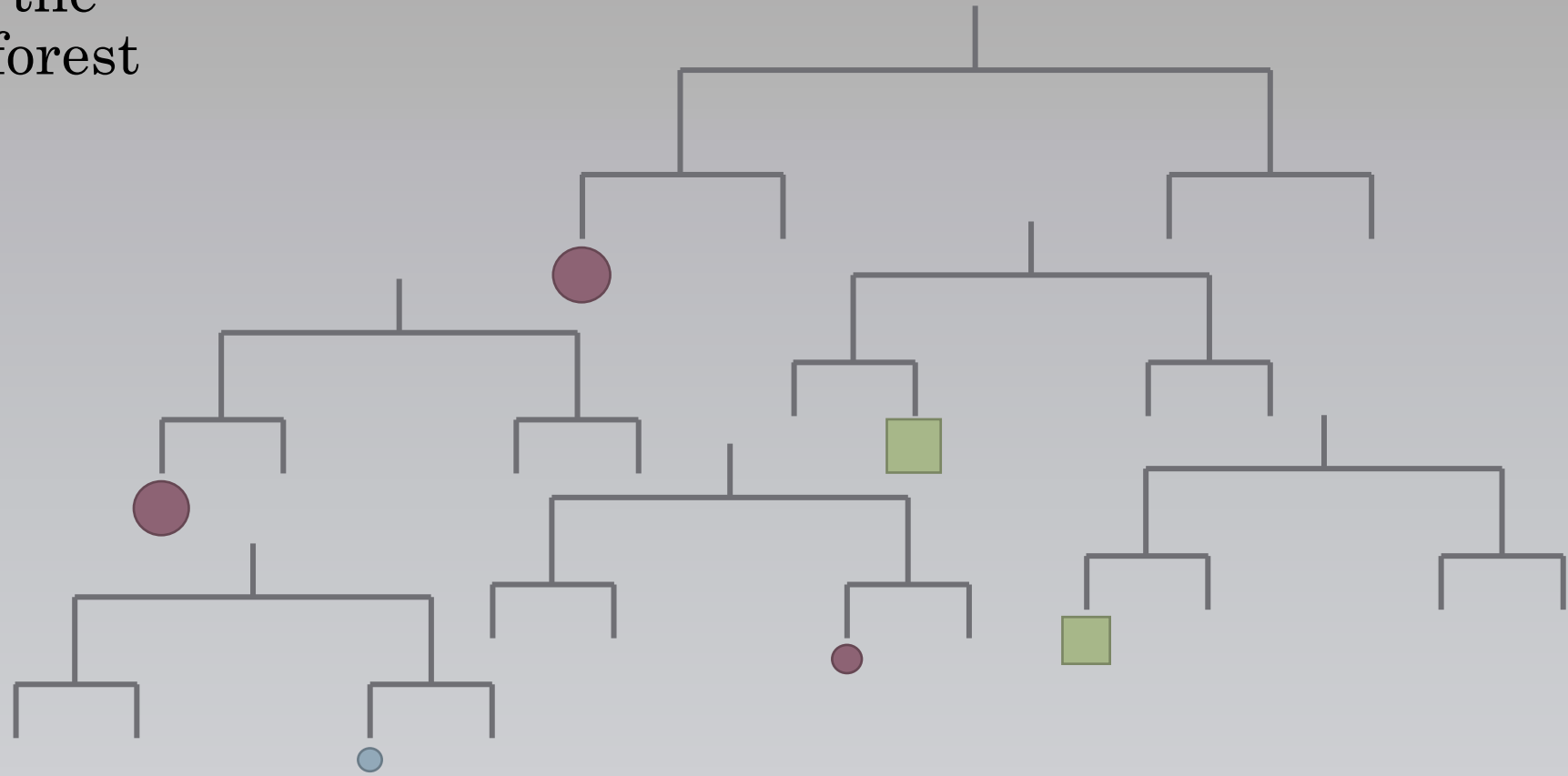
Follows the same procedure
but instead of using one tree,
it grows a forest of trees



2. Random Forest

And identifies based on the *consensus* of the whole forest

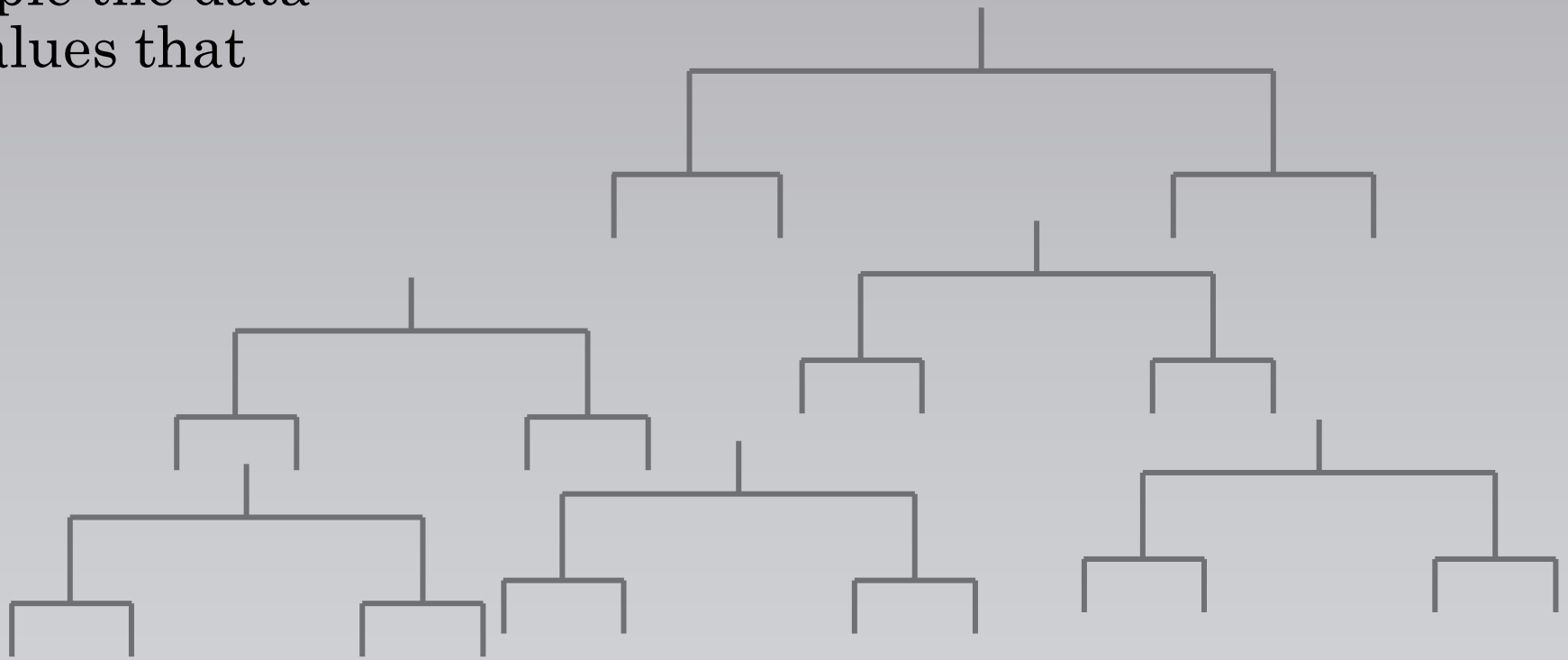
ID	Final Count
Red Circle	3
Blue Circle	1
Green Square	2

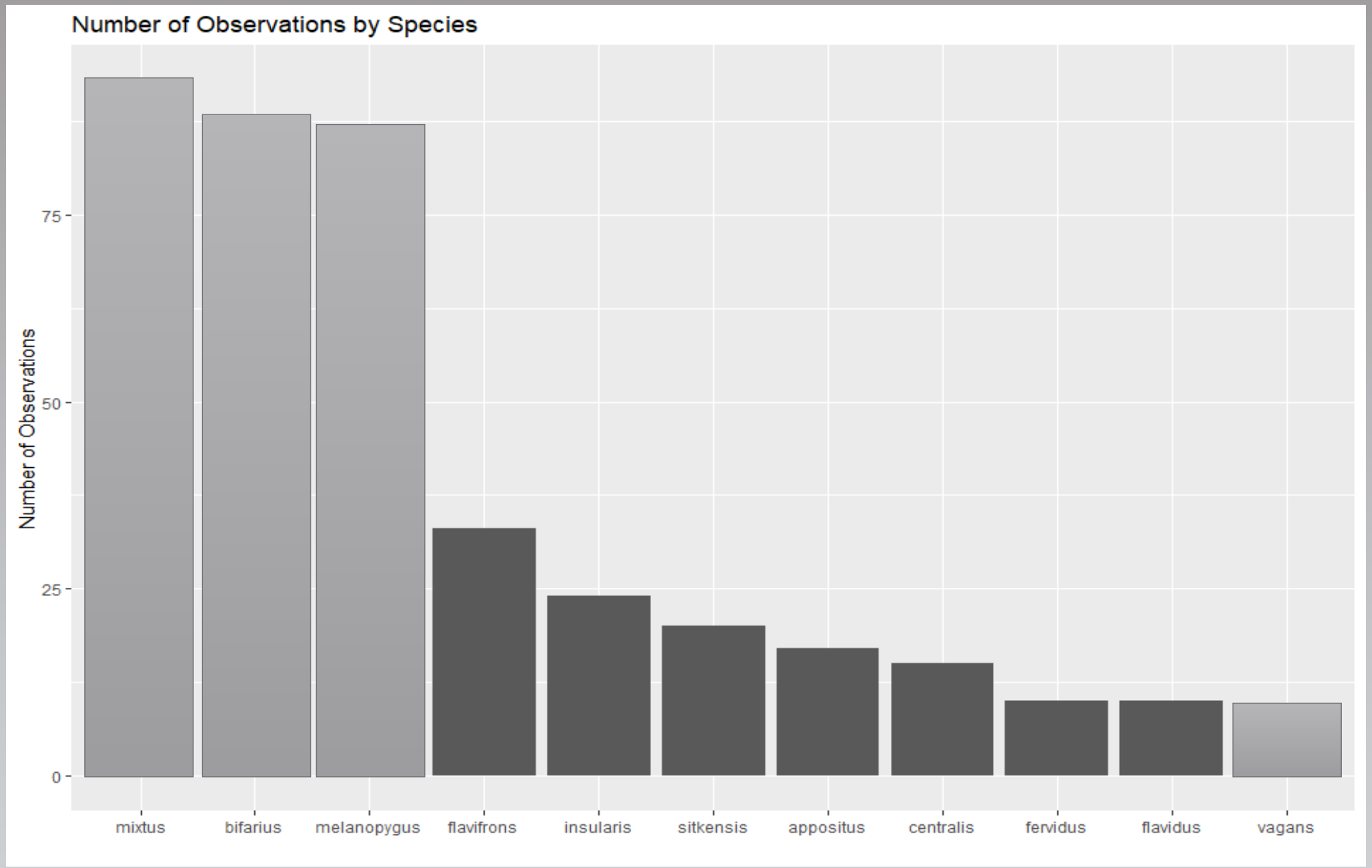


2. Random Forest

It also samples data with replacement.

This means it will sample the data more often, even the values that aren't very common.

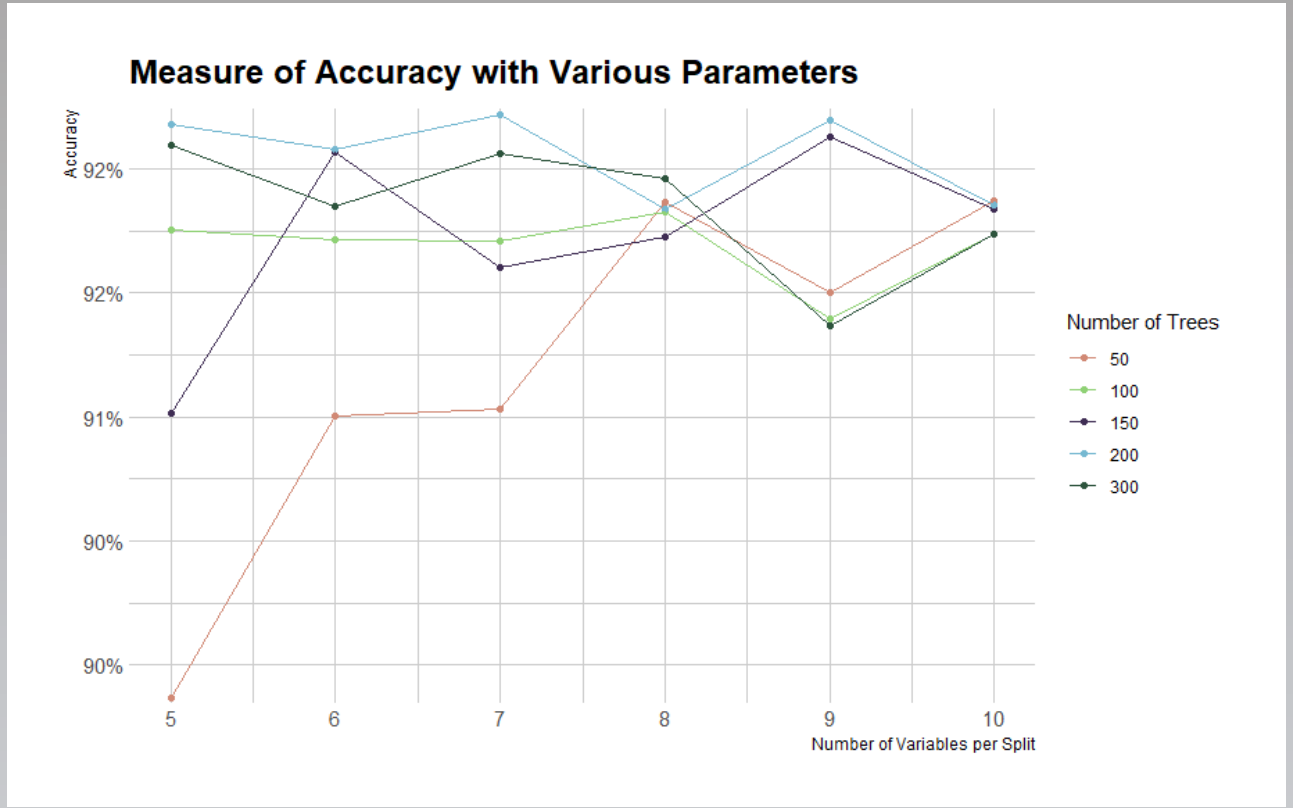


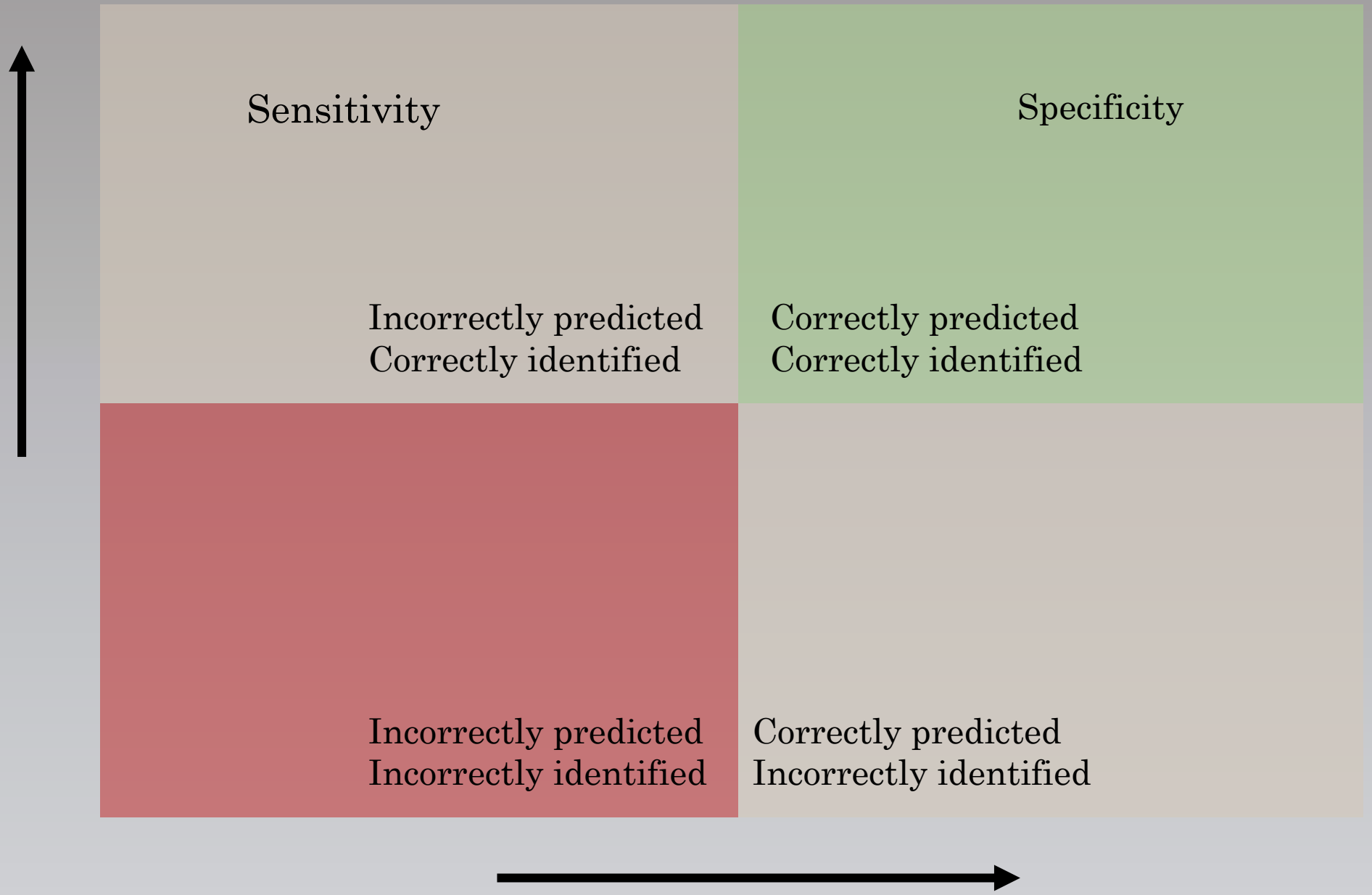

Number of Observations by Species
Number of Observations
75
50
25
0
mixtus
bifarius
melanopygus
flavifrons
insularis
sitkensis
appositus
centralis
fervidus
flavidus
vagans

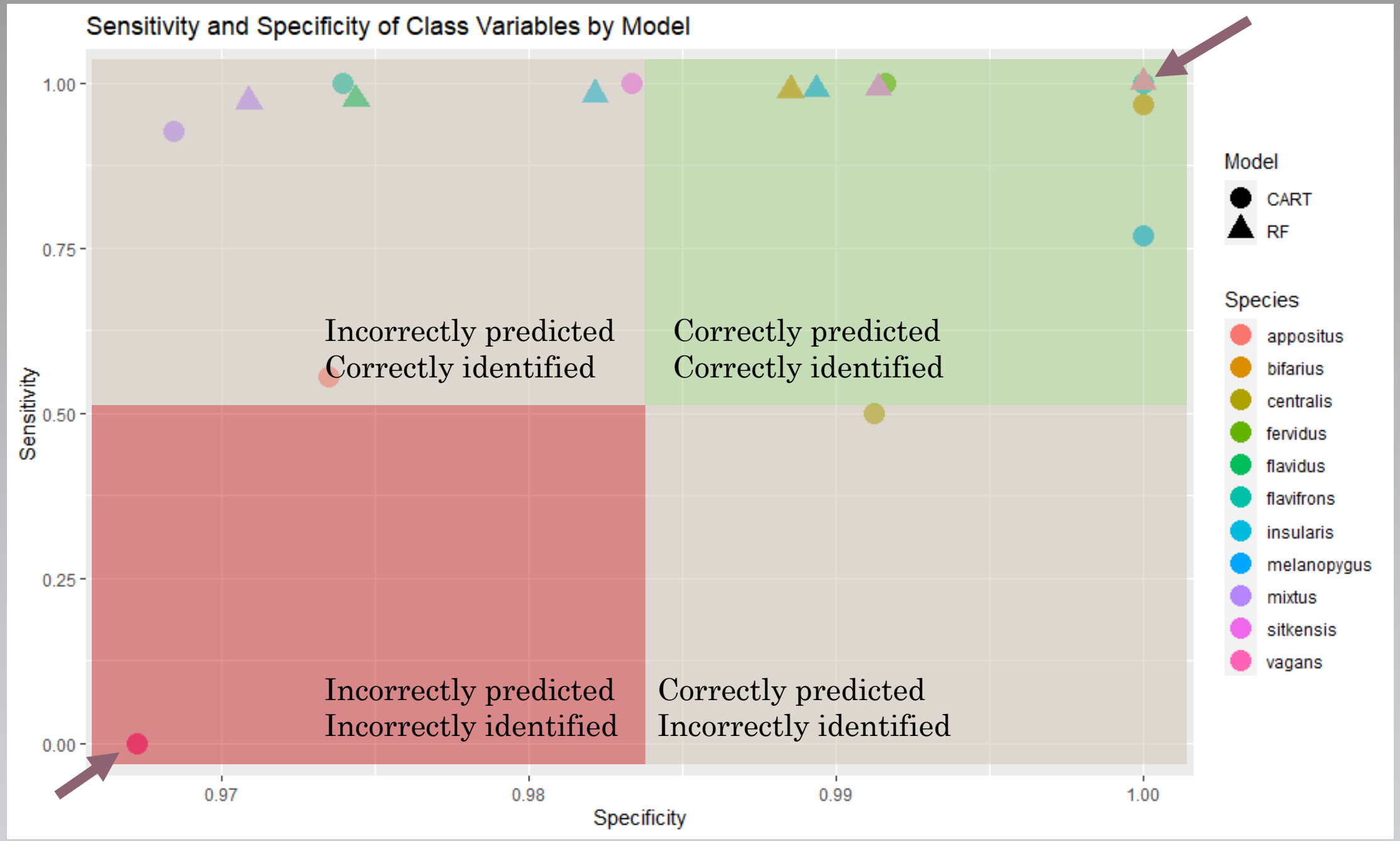
Pruning

- Models have to be “taught” how to look at the data.
- This involves analysis of
  - Accuracy
  - Error rates
  - Tests

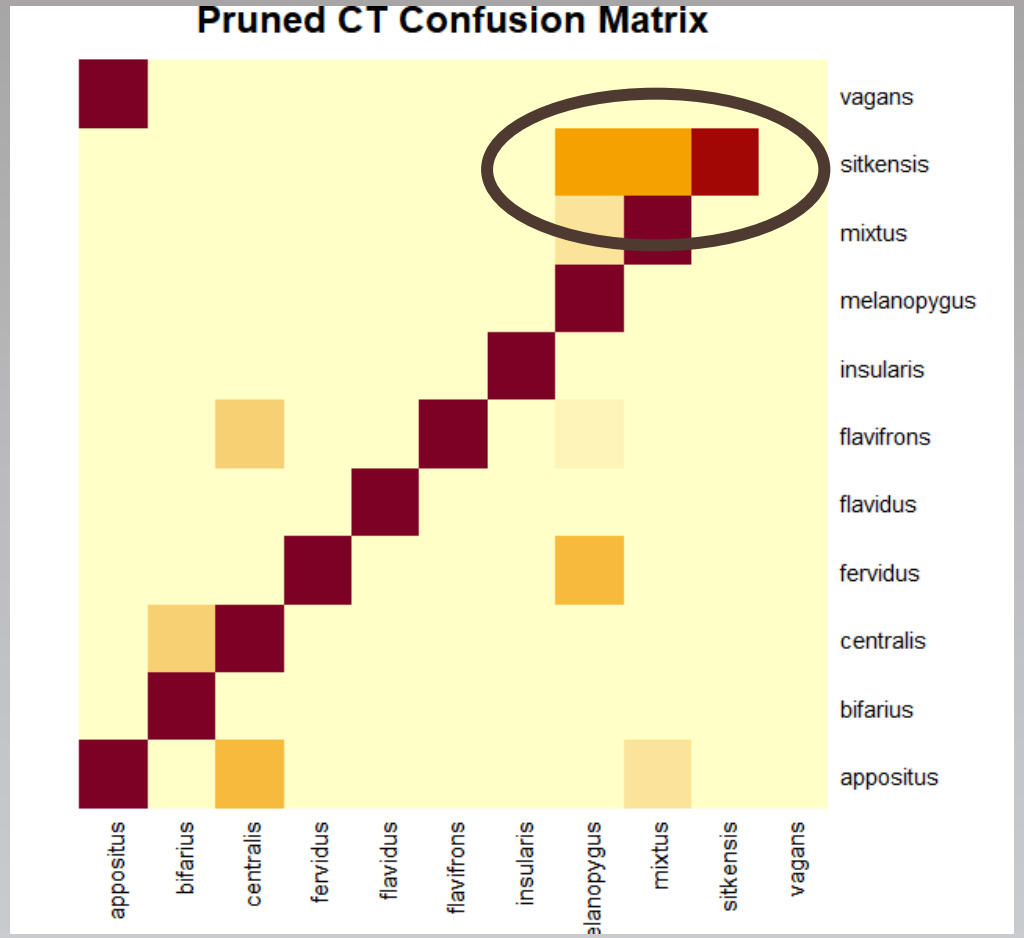
CART: 80% ⟶ 88%
RF: 90% ⟶ 93%



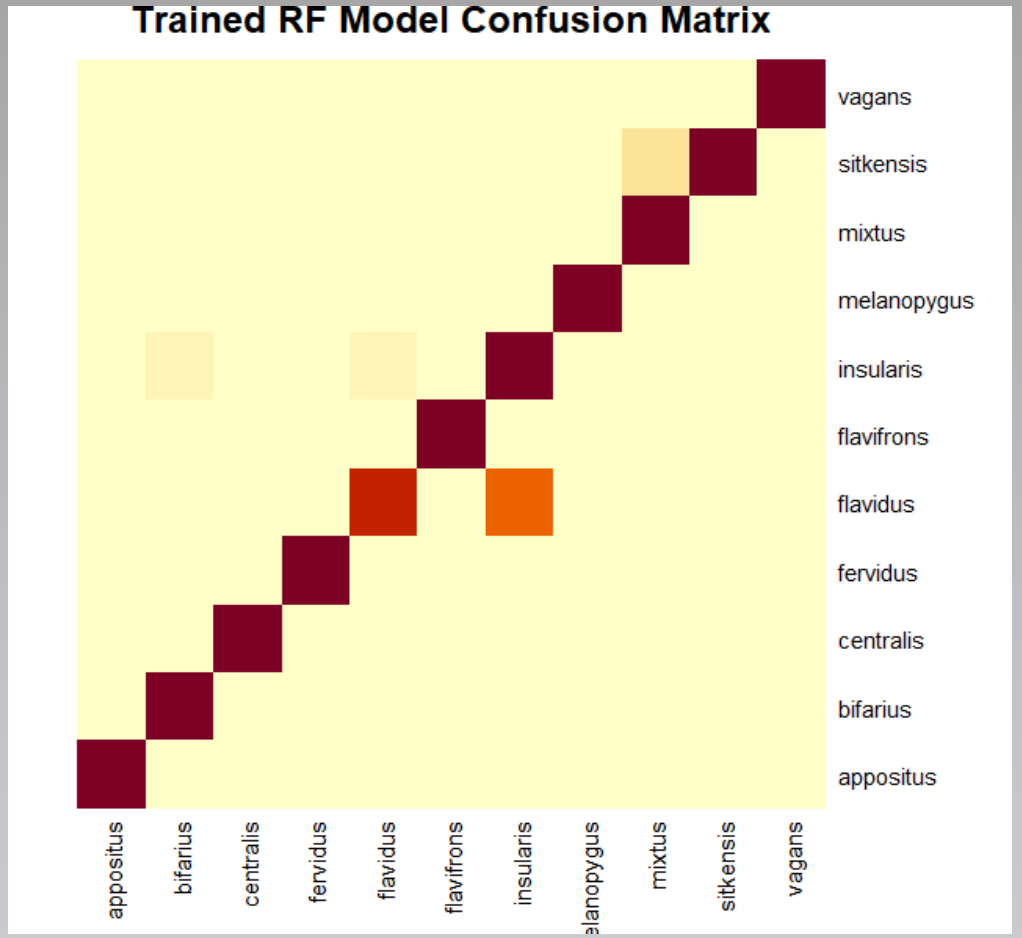

Sensitivity
Specificity
Incorrectly predicted
Correctly identified
Correctly predicted
Correctly identified
Incorrectly predicted
Incorrectly identified
Correctly predicted
Incorrectly identified


Sensitivity and Specificity of Class Variables by Model
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.97
0.98
0.99
1.00
Specificity
Incorrectly predicted
Correctly identified
Correctly predicted
Correctly identified
Incorrectly predicted
Incorrectly identified
Correctly predicted
Incorrectly identified
Model
CART
RF
Species
appositus
bifarius
centralis
fervidus
flavidus
flavifrons
insularis
melanopygus
mixtus
sitkensis
vagans

CART Model



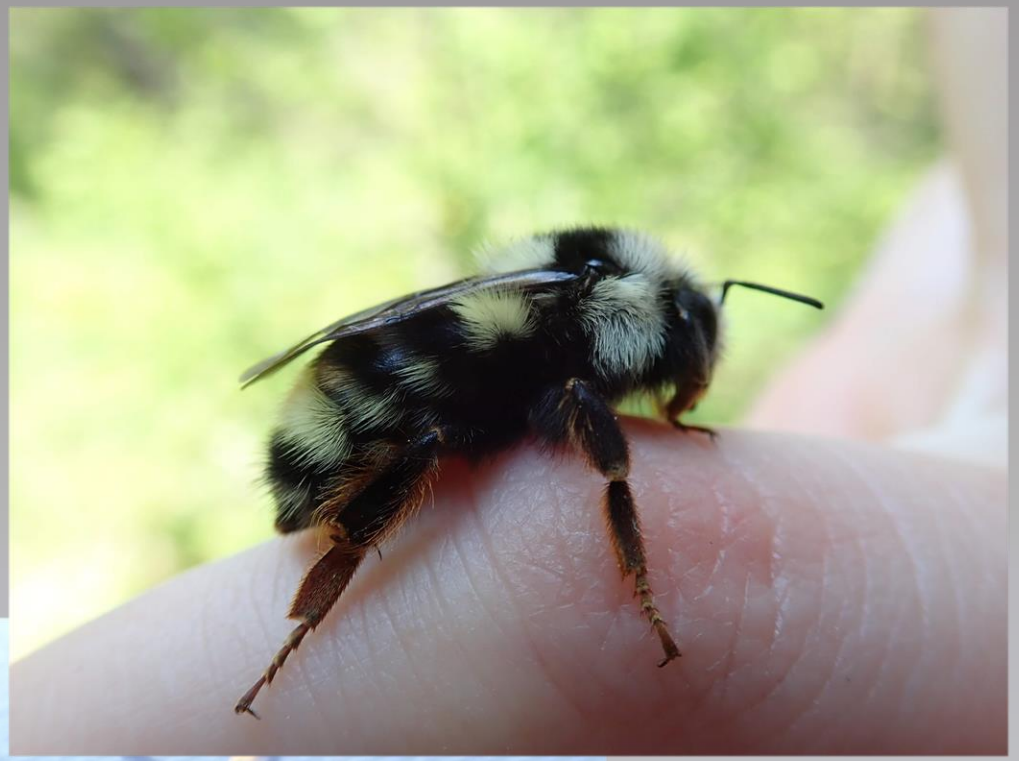
Random Forest Model

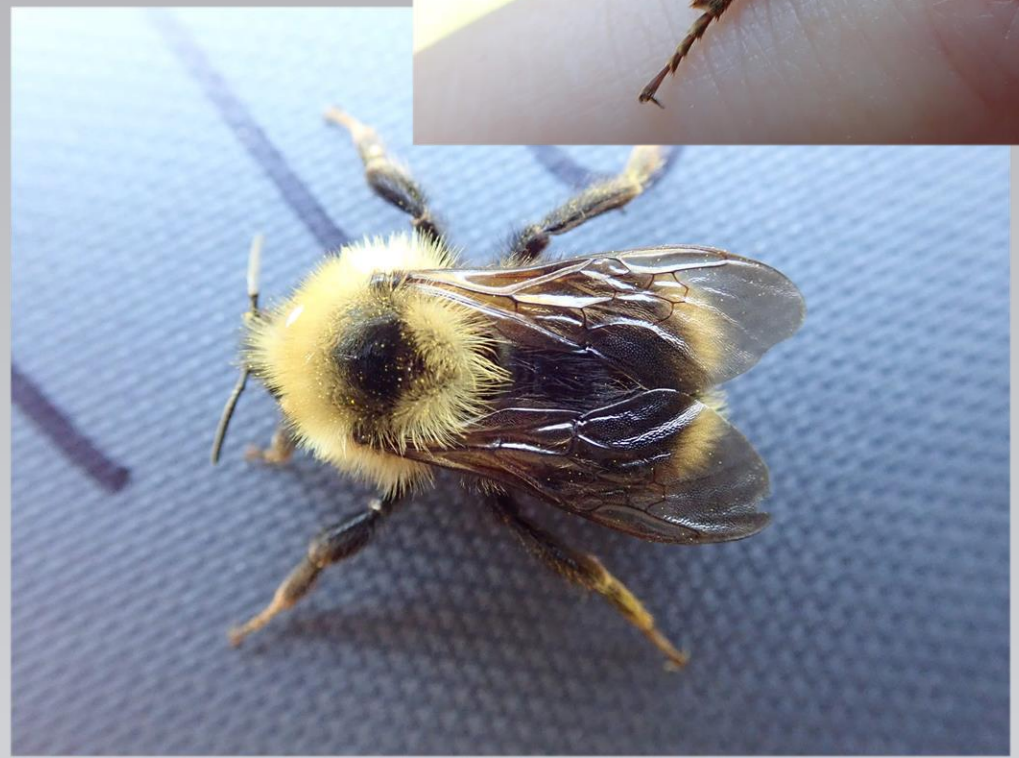


What does it mean?

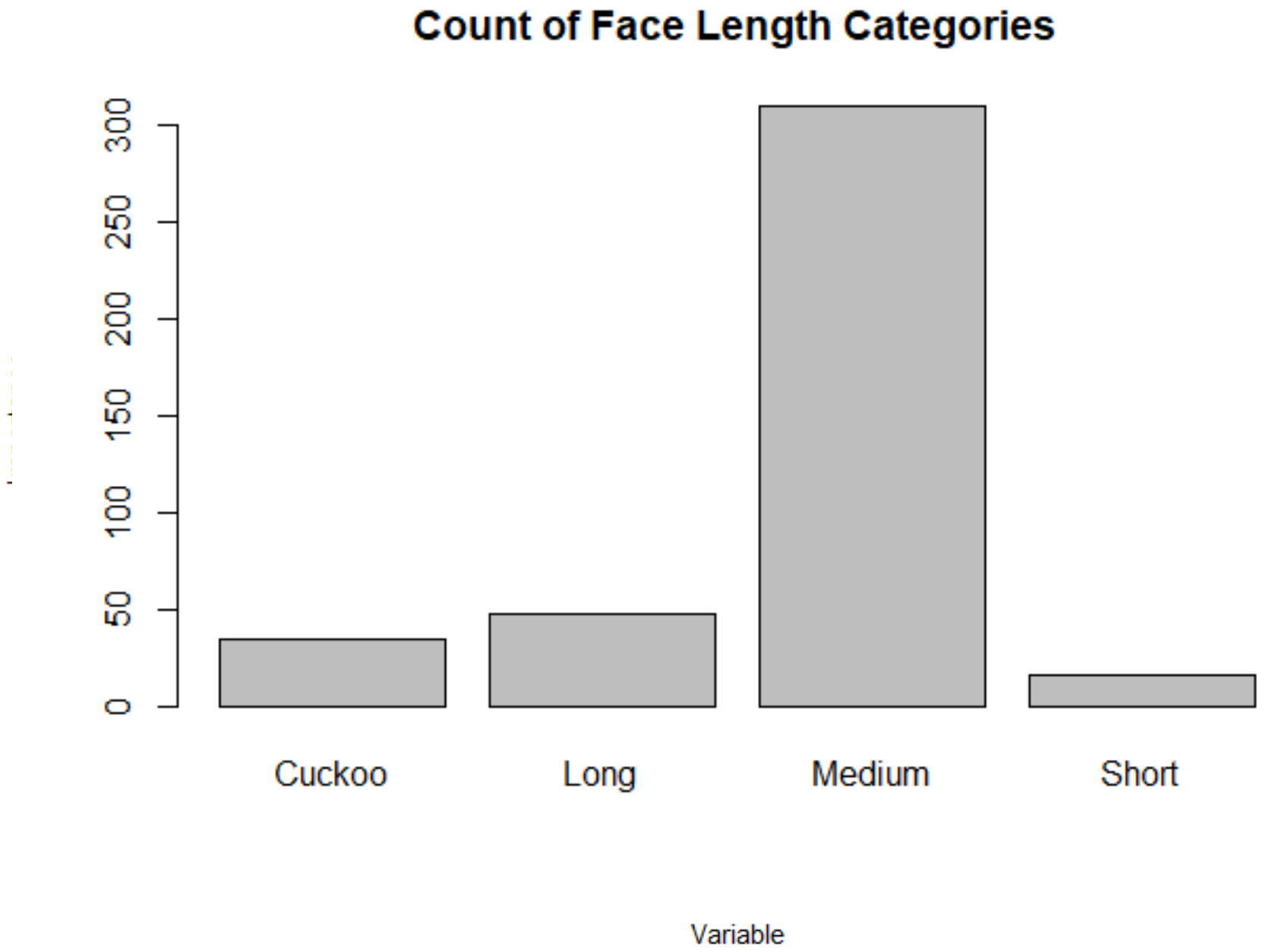
Random Forest model was more accurate and better at identifying species with fewer observations

The CART model was more nuanced in its identifications, even though it was worse at identifying species with fewer observations



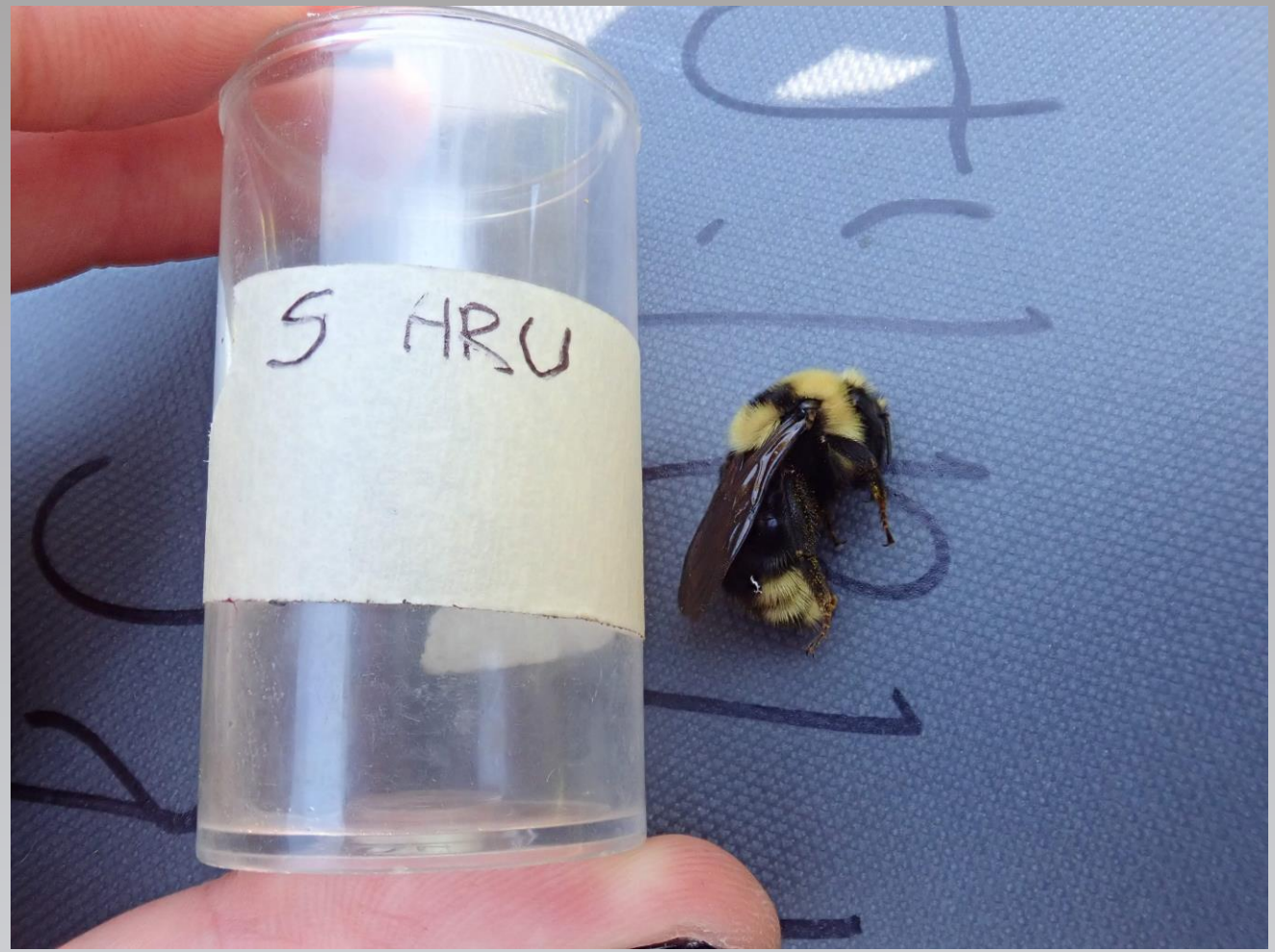


Variable Importance of Final Model



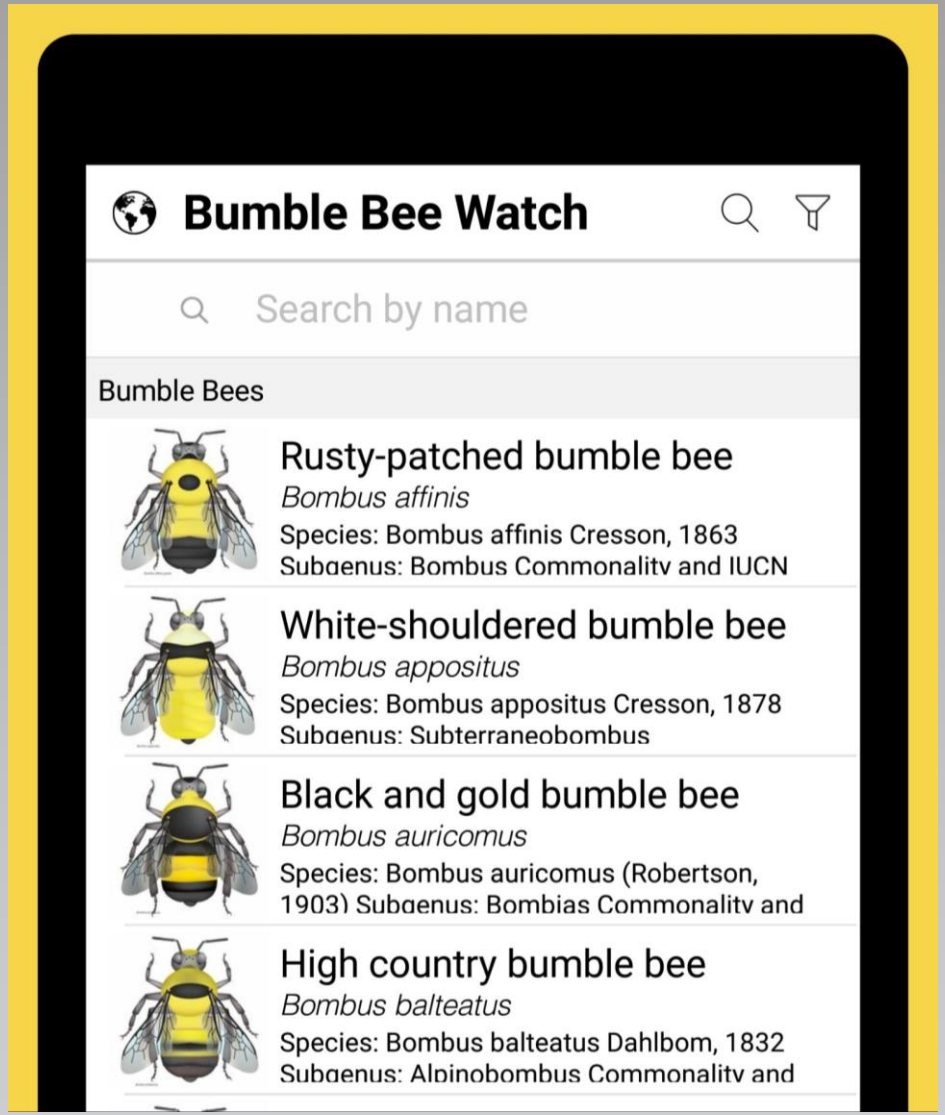
Future goals

- More data!
- Combine the nuance of CART with the accuracy of the RF
- Expand identifying factors



Why does it matter?

- With bumble bee populations declining, monitoring programs are becoming increasingly important.
- Community Science and the future of data collection
- Model frameworks can be applied to other species

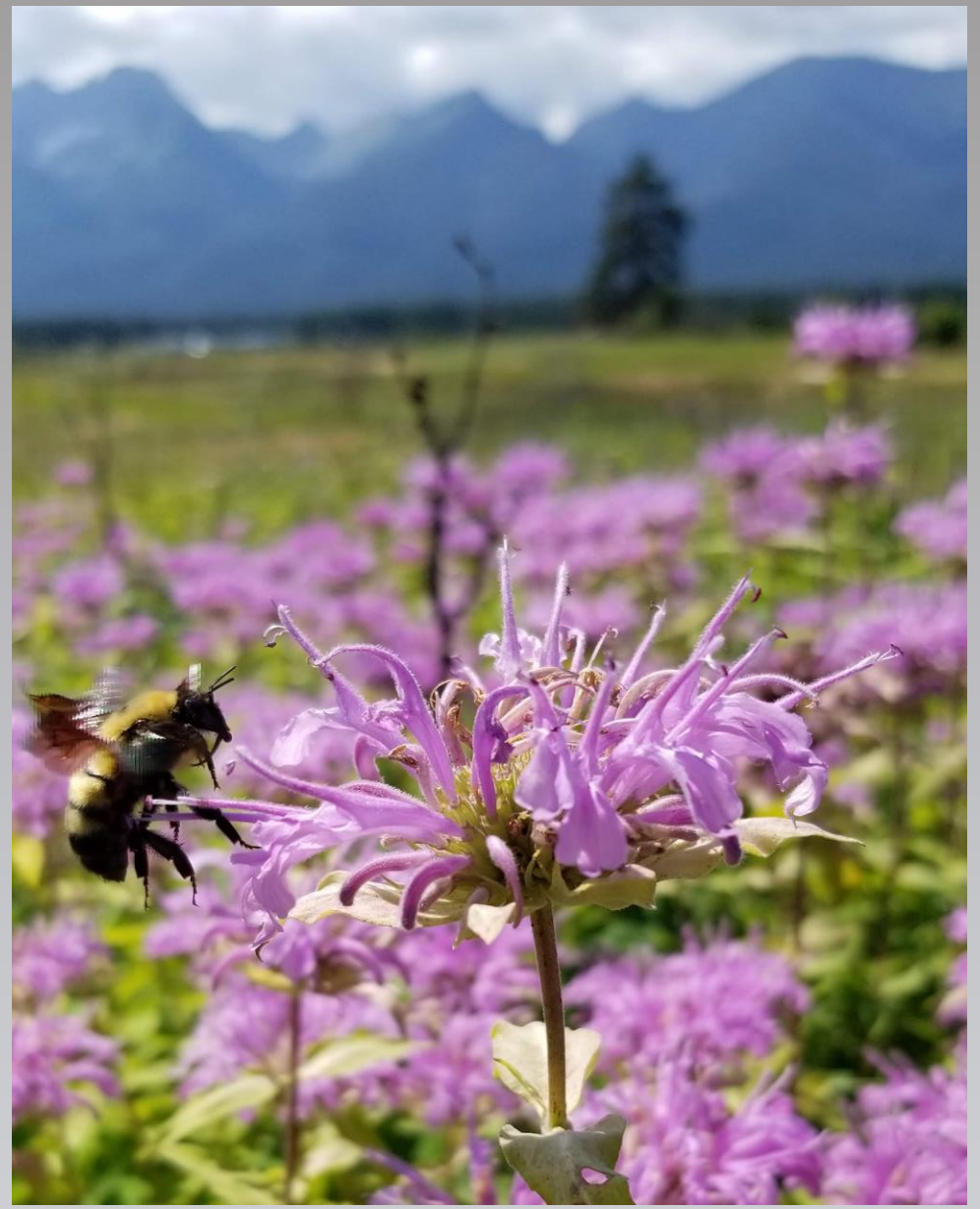


Citations

Rstudio Libraries used:

Caret	randomForest
Ggplot2	Rpart
Rpart.plot	E1071
Tidyverse	Ggally
Viridis	Parsnip
Dplyr	Hrbrthemes

Pictures: Salish Kootenai College Fisheries and Wildlife, Rebekah Brassfield, Bumble Bee Watch
Data property of Salish Kootenai College
Cameron, S., Lozier, J., Strange, J., Koch, J., Cordes, N., Solter, L., & Griswold, T. (2010, October). *Patterns of Widespread decline in North American Bumble Bees*. https://www.pnas.org/content/pnas/108/2/662.full.pdf
Acknowledgements:
Janene Lichtenberg, SKC




Thank you!